Title

Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell-of-origin model for prostate cancer heterogeneity.

Journal

Nature cell biology, 15(3)

ISSN

1465-7392

Authors

Wang, Zhu A
Mitrofanova, Antonina
Bergren, Sarah K
et al.

Publication Date

2013-03-01

DOI

10.1038/ncb2697

Peer reviewed

HHS Public Access

Author manuscript

Nat Cell Biol. Author manuscript; available in PMC 2013 September 01.

Published in final edited form as:

Nat Cell Biol. 2013 March ; 15(3): 274–283. doi:10.1038/ncb2697.

Lineage analysis of basal epithelial cells reveals their unexpected plasticity and supports a cell of origin model for prostate cancer heterogeneity

Zhu A. Wang[1,4], **Antonina Mitrofanova**[2,4], **Sarah K. Bergren**[1,4], **Cory Abate-Shen**[3,4], **Robert D. Cardiff**[5], **Andrea Califano**[2,4], and **Michael M. Shen**[1,4,*]

[1]Departments of Medicine and Genetics and Development, Columbia University College of Physicians and Surgeons, New York, NY 10032

[2]Department of Biomedical Informatics and Center for Computational Biology and Bioinformatics, Columbia University College of Physicians and Surgeons, New York, NY 10032

[3]Departments of Urology and Pathology and Cell Biology, Columbia University College of Physicians and Surgeons, New York, NY 10032

[4]Herbert Irving Comprehensive Cancer Center, Columbia University College of Physicians and Surgeons, New York, NY 10032

[5]Center for Comparative Medicine and Department of Pathology, School of Medicine, University of California, Davis 95616

Abstract

A key issue in cancer biology is whether oncogenic transformation of different cell types of origin within an adult tissue gives rise to distinct tumor subtypes that differ in their prognosis and/or treatment response. We now show that initiation of prostate tumors in basal or luminal epithelial cells in mouse models results in tumors with distinct molecular signatures that are predictive of human patient outcomes. Furthermore, our analysis of untransformed basal cells reveals an unexpected assay-dependence of their stem cell properties in sphere formation and transplantation assays versus genetic lineage-tracing during prostate regeneration and adult tissue homeostasis. Although oncogenic transformation of basal cells gives rise to tumors with luminal phenotypes, cross-species bioinformatic analyses indicate that luminal origin tumors are more aggressive than basal origin tumors, and identify a molecular signature associated with patient outcome. Our results reveal the inherent plasticity of basal cells, and support a model in which different cells of origin generate distinct molecular subtypes of prostate cancer.

*Author for correspondence at: phone: (212) 851-4723; fax: (212) 851-4572; mshen@columbia.edu.

Author contributions

Z.A.W. carried out mouse experiments, Z.A.W. and S.K.B. performed transplantation experiments, A.M. performed bioinformatic analyses, and R.D.C. analyzed tumor histology. Z.A.W. and M.M.S. designed the overall study, C.A.-S., R.D.C., A.C., and M.M.S. supervised the data analysis, and Z.A.W., A.M., and M.M.S. wrote the manuscript, in consultation with C.A.-S. and A.C. All authors provided discussion and comments on the manuscript.

The analysis of tumor cell of origin requires a detailed understanding of tissue cell types and their position in the lineage hierarchy[1]. In particular, stem cells are often considered to be excellent candidate cells of origin for cancer, given their inherent ability to self-renew. In the prostate gland, the three epithelial cell types are luminal cells, which express cytokeratins (CK) 8 and 18, and high levels of androgen receptor, basal cells, which express p63, CK5, and CK14, and rare neuroendocrine cells; in addition, a minor basal subpopulation known as "intermediate cells" co-express basal and luminal markers[2]. Notably, the adult prostate can undergo cycles of regression and regeneration following androgen ablation and restoration, implying that the prostate epithelium contains stem cells that function to promote regeneration.

To date, prostate stem cell populations have been identified in both the basal and luminal layers[3–7]. In particular, subpopulations of basal cells isolated using cell-surface markers display multipotency and self-renewal in sphere formation as well as tissue reconstitution assays[8–13]. Other work has identified a rare luminal population of castration-resistant Nkx3.1-expressing cells (CARNs) that displays stem cell properties in genetic lineage-tracing and tissue reconstitution assays[14]. It has been unclear whether these findings are mutually consistent, given the distinct assays for stem cell properties that have been employed.

The cell of origin model for intertumor heterogeneity proposes that tumor initiation from distinct cell types in the lineage hierarchy gives rise to tumor subtypes with different prognoses and/or treatment responses[1, 15]. Although this model has received considerable support in studies of breast cancer[16], it has not been systematically investigated in prostate cancer. However, several groups have investigated whether luminal cells or basal cells, or both, might serve as cell types of origin for prostate cancer. In particular, lineage-tracing analyses of CARNs have provided evidence that rare luminal cells can act as a cell of origin *in vivo*[14], while other studies have shown that lentiviral overexpression of oncogenes in isolated mouse and human basal cells can give rise to tumors with luminal phenotypes in renal grafts, whereas luminal cells fail to generate tumors under these conditions[13, 17]. In addition, a recent study has shown that both luminal and basal cells can serve as cells of origin for prostate cancer, generating tumors that are histologically similar in mouse models[18].

These previous studies have raised the possibility that *ex vivo* cell culture and tissue grafting assays may yield different results from *in vivo* lineage-tracing analyses. Therefore, we have undertaken a comprehensive analysis of prostate basal cell properties using genetic lineage-marking to examine the properties of the identical cell population in multiple assays for stem cell function. Our results show that apparent discrepancies in the published literature can be explained by the considerable plasticity of basal cells in distinct functional assays. Moreover, although both basal and luminal cells can serve as cells of origin for prostate cancer, giving rise to tumors with similar histological phenotypes, our molecular and bioinformatic analysis shows that the luminal origin tumors are more aggressive, and identifies a molecular signature that has predictive value for human patient survival. Thus, our study supports the cell of origin model as a basis for distinct prostate cancer subtypes.

Results

Analysis of lineage-marked prostate basal cells in cell culture and grafting assays

To provide a comprehensive analysis, we have performed genetic marking of prostate epithelial basal cells using a *CK5-CreER^{T2}* transgenic line[19] in combination with the *R26R-YFP* reporter allele[20] for isolation of a purified cell population for sphere formation and tissue reconstitution assays and for lineage-tracing *in vivo*. In control experiments, tamoxifen induction of hormonally-intact *CK5-CreER^{T2}; R26R-YFP/+* mice resulted in highly-specific expression of YFP in 24.5% (n=1,538/6,267) of CK5-positive basal cells in the anterior prostate lobe, while no YFP-positivity was observed in non-basal cells (n=0/15,846) (Fig. 1a); quantitation for all experiments is detailed in Supplementary Table S1. We verified that the YFP-marked cells were positive for the basal cell marker p63 and CK14, and were mostly negative for the luminal marker CK18 (Supplementary Fig. S1a,e–n), and were detected at similar frequencies in the dorsolateral (23.2%) and ventral (24.9%) prostate lobes (Supplementary Fig. S1c,d). Furthermore, 1.6% of p63^{+}YFP^{+} cells also expressed the luminal marker CK18 (Supplementary Fig. S1o–s), indicating that the marked population includes intermediate cells.

Next, we isolated lineage-marked YFP-positive cells by flow-sorting of dissociated prostate cells from tamoxifen-induced *CK5-CreER^{T2}; R26R-YFP/+* mice. Approximately 3.2% of total prostate cells were isolated in the YFP^{+} fraction (Fig. 1b), and greater than 98% of these isolated YFP^{+} cells were CK5-positive, p63-positive, and CK18-negative (Supplementary Fig. S2a–l). Furthermore, we compared this YFP^{+} population to that of Lin^{-}Sca-1^{+}CD49f^{hi} (LSC) cells, which have been previously characterized as a basal population enriched for stem/progenitor cells[8, 12]. Nearly all YFP^{+} cells (98.7%) were contained in the Lin^{-}CD49f^{+} population, and 8.0% of the YFP^{+} cells were present in the Lin^{-}Sca-1^{+}CD49f^{hi} fraction (Fig. 1c). Conversely, we found that 24.4% of the Lin^{-}Sca-1^{+}CD49f^{hi} cells from tamoxifen-induced *CK5-CreER^{T2}; R26R-YFP/+* mice were YFP^{+} (Fig. 1d,e), similar to the overall percentage of YFP^{+} basal cells. These results indicate that the YFP^{+} population includes Lin^{-}Sca-1^{+}CD49f^{hi} cells in a proportional and unbiased manner.

Using unsorted dissociated cells from tamoxifen-induced *CK5-CreER^{T2}; R26R-YFP/+* prostate tissue in sphere formation assays, we found that approximately 23% of the resulting spheres were YFP-positive (Fig. 1f), consistent with the overall percentage of marked basal cells. After flow-sorting, 4.7% of the YFP-positive cells could form spheres that displayed expression of basal and luminal markers (Fig. 1g; Supplementary Fig. S2m–r), consistent with previous studies[8, 21]. This frequency of sphere formation was similar between the Lin^{-}Sca-1^{+}CD49f^{hi} YFP-positive cells and the remaining YFP-positive cells (Supplementary Fig. S2s), and was similar in mice at both 2 months and 12 months of age (Fig. 1g).

Tissue reconstitution assays have been used to show that Lin^{-}Sca-1^{+}CD49f^{+} and Lin^{-}Sca-1^{+}CD49f^{hi}Trop2^{hi} cells readily generate prostate ducts in renal grafts[8, 12], but the overall efficiency of basal cells in this assay has not been previously determined. Therefore, we have used our highly purified YFP^{+} population in prostate reconstitution assays, and

have performed a limiting dilution analysis to determine the percentage of basal cells with reconstituting ability. This analysis shows that approximately 3.9% of basal cells contain graft-forming activity (Fig. 1h,i), which display their donor origin by expression of YFP, prostatic histology, and the formation of ducts with luminal, basal, and neuroendocrine cells (Fig. 1j–o; Supplementary Fig. S2t). Notably, the similar efficiencies of tissue reconstitution (3.9%) and sphere formation (4.7%) raises the possibility that these two distinct assays may identify the same progenitor population.

Identification of rare bipotential basal cells during prostate regeneration *in vivo*

Using lineage-tracing, we explored the properties of basal cells during prostate regeneration *in vivo*, using a protocol similar to one that we have previously employed[14]. We marked prostate basal cells by tamoxifen treatment of hormonally-intact *CK5-CreER^{T2}; R26R-YFP/+* mice, as in Fig. 1a, followed by androgen-deprivation to induce prostate regression, and then androgen restoration to promote prostate regeneration (Fig. 2a). After regeneration, we found that most YFP$^+$ cells were basal, but 0.04% of YFP$^+$ cells were luminal (n=5/11,427 cells) (Fig. 2b–f; Supplementary Table S1). Notably, the percentage of basal cells displaying bipotentiality as detected by lineage-tracing *in vivo* is significantly lower than the percentage displaying stem cell properties in sphere formation or tissue reconstitution assays ($p<0.0001$; χ^2 test).

We next asked whether this low detection rate of marked luminal cells might reflect an inherently low proliferation rate of basal cells (Fig. 2g,h; Supplementary Fig. S3). Using both Ki67 immunostaining and BrdU incorporation methods, we found that approximately 8.0% of basal cells proliferated during prostate regeneration (Fig. 2g,h; Supplementary Fig. S3). Furthermore, rare bipotential basal cells were observed at approximately 0.05% frequency using alternative regeneration protocols or in aged mice (Supplementary Fig. S4). Taken together, these results indicate that the frequency of luminal cell generation by dividing basal cells is low but stable during regeneration *in vivo* (Supplementary Table S1).

Given these findings, we then investigated whether lineage-marked luminal progeny of basal cells would accumulate during serial rounds of androgen-deprivation and androgen-restoration to drive prostate regression and regeneration. We analyzed the frequency of YFP$^+$ luminal cells among total YFP$^+$ cells after three and five rounds of serial regeneration, as well as after four rounds in the regressed state (Fig. 2i–o). We observed 0.6% marked luminal cells after three rounds and 3.4% after five rounds (Fig. 2j–l,n,o). Thus, there is a gradual increase in the percentage of luminal cells among YFP$^+$ cells with increasing rounds of regeneration, which is further supported by the detection of castration-resistant YFP$^+$ luminal cells in the regressed prostate after four rounds (Fig. 2m). Taken together, these results are consistent with a model in which a bipotential basal progenitor can give rise to luminal progeny with transit-amplifying characteristics.

Identification of rare bipotential basal cells during prostate homeostasis *in vivo*

We also examined the generation of luminal cells by lineage-marked basal cells during prostate homeostasis, using tamoxifen induction at 2 months of age followed by a long chase period (Fig. 3a). Similar to our analysis of serial regeneration, we found that the frequency

of YFP^+ luminal cells among all YFP^+ cells increased with time, as determined at 4 months (0.02%), 6 months (0.5%), or 12 months (3.0%) of age (Fig. 3b–f). However, using BrdU incorporation assays, we found that the proliferation rate in both basal (1.1%) and luminal (6.1%) compartments was relatively stable at 2, 6, and 12 months of age (Fig. 3g–n). These results suggest that low-frequency generation of luminal cells from a bipotential basal progenitor results in small clusters of lineage-marked luminal cells, due to the higher proliferation rate of luminal cells during homeostasis. Furthermore, these findings indicate that basal cells display bipotentiality during regeneration and tissue homeostasis at approximately similar frequencies.

Basal cells as a cell of origin for prostate cancer

Next, we investigated whether $CK5^+$ basal cells can be a cell of origin for prostate cancer, using a conditional allele of the *Pten* tumor suppressor for inducible inactivation in mice, modeling one of the most frequent genetic alterations in human prostate cancer[22, 23]. We found that *CK5-CreER^{T2}; Pten$^{flox/flox}$; R26R-YFP/+* prostates at 1 month following tamoxifen induction displayed small foci of epithelial hyperplasia and Grade I PIN lesions[24] in otherwise histologically normal glands (Fig. 4a,b; Supplementary Table S2). At this stage, we could detect phosphorylated Akt in basal cells, as expected after loss of *Pten* (Fig. 4f). By 3 months after induction, these mice displayed Grade II and III PIN lesions, while at 6 months they had Grade III and IV lesions (Fig. 4c,d; Supplementary Fig. S5; Supplementary Table S2). Interestingly, we observed increased cell proliferation in basal cells prior to PIN lesion formation (Fig. 4g; Supplementary Table S2), and subsequently in the PIN/tumor lesions (Fig. 4h). Notably, even small PIN lesions contained few $CK5^+$ basal cells (Fig. 4i,m), but instead were mostly comprised of $CK18^+$ and AR^+ luminal cells (Fig. 4j,k,m), indicating that oncogenic transformation of basal cells promotes luminal differentiation of their progeny. Furthermore, the low frequency of $CK5^+CK18^+$ intermediate cells in the untransformed prostate (1.6%; Supplementary Fig. 1o–s) and during tumor formation (3.0%; Fig. 4l,m) suggests that luminal differentiation occurs from transformed basal cells, and does not arise exclusively from intermediate cells.

We next examined whether the time course and histopathology of PIN/tumor lesions in *CK5-CreER^{T2}; Pten$^{flox/flox}$; R26R-YFP/+* mice differed from those in *Nkx3.1$^{CreERT2/+}$; Pten$^{flox/flox}$; R26R-YFP/+* mice (Fig. 4e; Supplementary Fig. S6a–n), which have a luminal cell of origin. Consistent with a recent report[18], we found that the overall histopathology of the luminal origin PIN/tumor lesions was very similar, but arose with a different time course. At 1 month after induction, the *Nkx3.1$^{CreERT2/+}$; Pten$^{flox/flox}$; R26R-YFP/+* mice displayed Grade II and III PIN, resembling 3-month basal origin lesions (Fig. 4c; Supplementary Fig. S6b; Supplementary Table S2). Similarly, 3-month luminal origin tumors displayed Grade III and IV PIN, resembling 6-month basal origin tumors (Fig. 4d,e; Supplementary Fig. S6c,d; Supplementary Table S2). These histopathological similarities were further supported by marker analyses (Fig. 4h–k; Supplementary Fig. S6f–k). However, we did note that $CK5^+CK18^+$ cells occurred at low frequencies in all prostate lobes of basal origin lesions, but at higher frequencies in the anterior and dorsolateral lobes of luminal origin lesions (Fig. 4l,m; Supplemental Fig. S6l–n). Overall, basal origin tumors are histologically similar to luminal origin tumors, but arise more slowly, perhaps due to

differences in the starting number of cells undergoing transformation and/or an intrinsic delay due to luminal differentiation from basal cells.

A luminal origin molecular signature predicts patient survival

To determine whether basal and luminal origin tumors might display molecular differences, we performed transcriptome analyses comparing PIN/tumor lesions from basal or luminal origins at time points at which they displayed similar histopathological phenotypes. We used RNA-seq to profile prostate tissue from basal origin tumors of *CK5-CreER^{T2}; Pten$^{flox/flox}$; R26R-YFP/+* mice that were uninduced (control), or at 3 or 6 months after tamoxifen induction, or from luminal origin tumors of *Nkx3.1$^{CreERT2/+}$; Pten$^{flox/flox}$; R26R-YFP/+* mice that were uninduced (control), or at 1 or 3 months after induction (6 mice/category). Principal Components Analysis demonstrated the reproducibility of the independent biological replicates (Fig. 5a,b). We then performed Gene Set Enrichment Analysis (GSEA) to compare "initiation" signatures for basal origin lesions (3 months vs. control) and for luminal origin lesions (1 month vs. control) (Supplementary Tables S3, S4); a similar comparison was performed for "progression" signatures for basal origin (6 months vs. 3 months) and luminal origin lesions (3 months vs. 1 month) (Supplementary Tables S5, S6). These comparisons demonstrated the strong reciprocal enrichment of the "initiation" as well as "progression" signatures (Fig. 5c,d), indicating that basal and luminal origin tumors are globally similar at the molecular level.

Nonetheless, we could identify molecular differences between the luminal and basal origin signatures by applying bioinformatic subtraction of the similar components of their signatures. Thus, we generated a mouse expression signature containing genes up-regulated in luminal origin lesions relative to basal origin lesions (luminal 3 months – basal 6 months), or conversely up-regulated in basal origin lesions (Supplementary Table S7). Next, we used a mouse-to-human cross species approach to compare this mouse luminal vs. basal signature to a human "lethality signature" generated using survival data in a Swedish "watchful-waiting" patient cohort[25]. We defined this lethality signature as a list of genes ranked by their differential expression between high-risk (death within 12 months, 6 samples) and low-risk (survival for more than 192 months, 12 samples) cases in this cohort (Supplementary Table S8). Using GSEA, we found that genes over-expressed in luminal origin tumors were significantly enriched in genes up-regulated in the lethality (high-risk) signature (Fig. 5e). In contrast, genes over-expressed in the basal origin tumors displayed no statistical enrichment in the human lethality signature (Fig. 5f), suggesting that luminal origin prostate cancer is more aggressive than basal origin prostate cancer. Furthermore, we used the luminal vs. basal origin signature to infer Gene Ontology-Biological Process (GO-BP) gene sets that were significantly up- or down-regulated. We identified 57 such GO-BP categories ($p<0.05$) (Fig. 5g; Supplementary Table S9), including categories such as "cell cycle mitotic", "E2F–mediated regulation of DNA replication", and "extension of telomeres", consistent with a more aggressive phenotype of luminal origin tumors.

Next, we investigated the clinical relevance of the 68 genes with significant contribution to the enrichment of the luminal origin up-regulated signature in the human lethality signature (the "leading edge" in Fig. 5e; Supplementary Table S10). Interestingly, many of these 68

genes are up-regulated in human prostate cancer, as shown by analysis of five human patient datasets[22, 26–29] (Supplementary Fig. S6o; Supplementary Table S10). We tested this luminal origin leading-edge signature (LOLES) with two different datasets from Memorial Sloan-Kettering Cancer Center (MSKCC)[22, 30], in which patients are characterized by their biochemical recurrence (BCR) free survival time, representing the duration between prostatectomy and subsequent detection of rising serum prostate-specific antigen (PSA) levels. Using *k*-means clustering, we could stratify the 79 primary tumor samples from the Glinsky dataset[30], and the 131 primary tumor samples from the Taylor dataset[22], into two groups with significant differences in BCR-free survival (log-rank p=0.0255, Fig. 5h; log-rank p=0.0245, Fig. 5i). Furthermore, the LOLES successfully stratified 263 patients of the Swedish cohort (excluding the samples used to define the lethality signature) into two distinct groups with a statistically significant difference in mean survival time of 3.4 years at 50% survival (log-rank p=0.0065, Fig. 5j).

Finally, we examined whether the LOLES may have independent prognostic value compared to histological Gleason scoring, which remains the best prognostic marker for overall survival[31]. Indeed, C-statistics analysis using the Swedish watchful-waiting cohort revealed that the LOLES improves the prognostic value of Gleason score from 0.76 (C = 0.76; 95% CI 0.72–0.80, p=1.10e–39) to 0.82 (C=0.82, 95% CI 0.77–0.84, p=2.25e–80) (Fig. 5k). Taken together, our bioinformatic analysis shows that the LOLES is highly correlated with poor patient prognosis.

Discussion

Using a lineage-marking approach to examine the same population in *ex vivo* and *in vivo* assays, we have shown that prostate basal cells display distinct properties in different assays for stem cell function, and reconcile several aspects of previous work on prostate stem cells that have appeared to be discordant. Notably, studies using *ex vivo* cell culture and tissue reconstitution assays have identified prostate basal cells as stem cells[8, 12, 21], but our analyses indicate that basal cells can exhibit substantial plasticity when removed from their endogenous tissue microenvironment. More generally, our work supports the notion that genetic lineage-tracing *in vivo*, not transplantation-based assays, represents a “gold standard” for identification of physiologically-relevant stem cells[32], and that unexpected plasticity should be considered when interpreting the outcomes of other stem cell assays.

In principle, the plasticity of prostate basal cells might be regulated by extrinsic or intrinsic factors. For example, stromal and/or luminal cells could inhibit basal cell plasticity, while *Pten* loss might confer independence from this inhibition, allowing formation of transformed luminal cells. Moreover, the embryonic urogenital mesenchyme employed in tissue reconstitution assays has potent reprogramming activity[33, 34], and could perhaps reprogram adult prostate basal cells to an embryonic multipotent progenitor state. Another possibility is that basal cell plasticity may be intrinsically regulated by cellular proliferation, since the percentage of luminal progeny is approximately 1–2% of dividing basal cells during regeneration and adult homeostasis. Thus, in response to strong proliferative signals provided by embryonic urogenital mesenchyme in tissue reconstitution assays, or due to *Pten* inactivation, basal cells might generate increased numbers of luminal cells.

Our findings suggest an overall similarity in the functional properties of basal cells in the prostate to those in other ductal epithelial tissues. Recent lineage-tracing analyses of the mammary and sweat gland epithelium have shown that luminal and myoepithelial cells are maintained by distinct unipotent progenitors, while myoepithelial cells can generate luminal cells in fad pad transplantation assays[35, 36]. Our work, together with other recent studies[18, 37], indicates that a similar lineage relationship exists in the prostate epithelium, and that the plasticity of myoepithelial/basal cells may be a conserved attribute of ductal epithelial tissues. Moreover, the plasticity of basal cells in the adult prostate epithelium may reflect their role as stem/progenitor cells that generate all epithelial cell types during organogenesis[38], similar to basal progenitors in mammary and sweat gland development[35, 36].

We have also detected rare prostate basal cells that display bipotentiality during adult tissue regeneration and homeostasis, at a frequency similar to that of bipotential luminal CARNs in the regressed prostate[14]. This finding differs from the conclusions of a recent report that the prostate basal lineage is completely unipotent[18]. The basis for this discrepancy is unclear, but could be due to the different Cre drivers utilized, or the larger number of cells and rounds of regeneration analyzed in this study. Notably, the existence of bipotential adult basal stem/progenitor cells is also supported by retrospective lineage analyses of human prostate tissue using mitochondrial DNA mutations[39, 40]. Taken together, our findings suggest that basal and luminal lineages are largely independent in the adult prostate, but rare basal and luminal stem/progenitor cells can potentially compensate for imbalances in cell number during regeneration and tissue homeostasis. One possible model is that rare basal (and luminal) stem/progenitor cells may reside at the top of an epithelial lineage hierarchy, while a larger subpopulation of basal cells, perhaps corresponding to transit-amplifying cells, can display plasticity in sphere and tissue reconstitution assays (Fig. 6a). Alternatively, basal cells may display stochastic stem/progenitor properties, with a low probability in the adult prostate epithelium and a higher probability when explanted or transformed (Fig. 6b), resembling the maintenance of interfollicular and esophageal epithelium by stochastic progenitors during homeostasis and wound repair[41, 42].

In humans, prostate adenocarcinoma displays a strong luminal phenotype with relatively uniform histopathological characteristics, and lacks distinct histological subtypes. At the molecular level, there has been some success in classifying tumors on the basis of gene expression profiling[26, 43], while distinct molecular subtypes may be identifiable by specific mutations and/or chromosomal rearrangements[22, 23, 44, 45]. Nonetheless, prostate cancer has previously appeared to differ from other human cancers in which distinct tumor subtypes are readily defined[16, 46–48]. However, although we find that basal and luminal origin tumors in mice are histologically similar, as also reported by Choi and colleagues[18], our mouse-to-human cross-species bioinformatic analysis has identified a molecular signature in luminal origin tumors that correlates with patient outcome. Thus, our analysis suggests that prostate tumors arising from different cell types of origin may have distinct prognostic outcomes and/or treatment responses.

A major clinical challenge in prostate cancer research has been to distinguish the minority of patients who will develop aggressive disease from those who have indolent cancer and

require minimal treatment. Indeed, considerable variation in outcome can exist between tumors with identical Gleason scores[31], indicating that intertumor heterogeneity exists among prostate cancers that are histologically indistinguishable. To date, it has been difficult to identify useful prognostic biomarkers to improve upon Gleason scoring, despite considerable effort to identify molecular signatures that can successfully stratify aggressive from non-aggressive disease[25]. Our cross-species analyses suggest that the cell of origin may represent an important component in determining prostate cancer aggressiveness, and may therefore lead to the identification of successful biomarkers.

Methods

Mouse strains and genotyping

The *Nkx3.1*^{CreERT2} targeted allele[14] and *CK5-CreER*^{T2} transgenic line[19] have been described previously. Genotyping was performed by PCR using tail genomic DNA, with the following primer sequences: *Nkx3.1* wild-type allele, 5′-CTCCGCTACCCTAAGCATCC-3′ and 5′-GACACTGTCATATTACTTGGACC-3′; *CreER*^{T2} allele, 5′-CAGATGGCGCGGCAACACC-3′ and 5′-GCGCGGTCTGGCAGTAAAAAC-3′; *Pten*^{flox} allele, 5′-ACTCAAGGCAGGGATGAGC-3′ and 5′-GTCATCTTCACTTAGCCATTGG-3′; *R26R-YFP* allele, 5′-GCGAAGAGTTTGTCCTCAACC-3′ (mutated forward), 5′-GGAGCGGGAGAAATGGATATG-3′ (wild-type forward) and 5′-AAAGTCGCTCTGAGTTGTTAT-3′ (wild-type and mutated reverse).

Mouse procedures

Castration of adult male mice was performed using standard techniques, with the fully regressed state attained at 4 weeks after castration. For prostate regeneration, testosterone (Sigma) was dissolved at 25 mg/ml in 100% ethanol and diluted in PEG-400 to a final concentration of 7.5 mg/ml. Testosterone was administered for 4 weeks at a rate of 1.875 μg/h delivered by subcutaneous implantation of mini-osmotic pumps (Alzet). This regimen yields physiological levels of serum testosterone[49].

For tamoxifen induction, mice were administered 9 mg/40 g tamoxifen (Sigma) suspended in corn oil, or vehicle alone for negative controls, by oral gavage once daily for 4 consecutive days, followed by a chase period of 14 days. BrdU (100 mg/kg) (Sigma) was administered by intraperitoneal injection twice daily for 12 consecutive days during regeneration or homeostasis to label proliferating cells.

Tissue collection and flow cytometry

For histological and immunofluorescence analysis, individual prostate lobes or renal grafts were dissected and fixed in 4% paraformaldehyde for subsequent cryoembedding in OCT compound (Sakura), or fixed in 10% formalin followed by paraffin embedding. For RNA isolation and RNA-seq analysis, prostate tissues were quickly dissected, flash-frozen in liquid nitrogen, and stored at −80°C.

For flow cytometry, prostate tissues were dissected and minced to small clumps, followed by enzymatic dissociation with 0.2% collagenase I (Invitrogen) in DMEM media with 5%

FBS for 3 h at 37°C. Tissues were digested with 0.25% Trypsin-EDTA (StemCell Technologies) for 1 h at 4°C, passed through 21- to 26-gauge syringes and filtered through a 40-μm cell strainer to obtain single-cell suspensions. Dissociated prostate cells were suspended in Hanks' Balanced Salt Solution Modified/2% FBS. Cell sorting was performed on a BD FACS Aria II instrument in the Flow Cytometry Shared Resource of the Herbert Irving Comprehensive Cancer Center. Antibodies used for sorting of $Lin^{-}Sca\text{-}1^{+}CD49f^{hi}$ cells are listed in Supplementary Table S11.

Sphere formation assay

For sphere formation assays, dissociated cells were incubated in PrEGM medium (Lonza). For each sample, 40 μl of cell suspension was mixed with 60 μl cold Matrigel, and pipetted around the rim of a well of a 12-well plate. The plates were placed in a 37°C CO_2 incubator for 30 min to allow the Matrigel to solidify. 800 μl warm PrEGM was then added to each well. The spheres were cultured and monitored for 7–10 days with 50% medium change every 3 days. For sphere differentiation experiments, the spheres were subsequently cultured in PrEGM medium with 10^{-8} M dihydrotestosterone (DHT), and monitored for 6 days with 50% medium change every 2 days.

Tissue reconstitution assay

For limiting dilution analysis, 5000, 100, or 20 dissociated YFP^{+} cells obtained from tamoxifen-induced *CK5-Cre^{ERT2}; R26R-YFP/+* mice were mixed with 2.5×10^{5} dissociated urogenital sinus mesenchyme (UGM) cells from embryonic day 18.0 rat embryos. UGM cells were obtained from dissected urogenital sinus treated for 30 min in 1% trypsin, followed by mechanical dissociation and treatment with 0.1% collagenase B (Roche) for 30 min at 37°C, and washing in PBS. Pelleted cell mixtures were resuspended in 10 μl of 1:8 collagen:setting buffer (10× Earle's Balanced Salt Solution (Life Technologies), 0.2 M $NaHCO_3$ and 50 mM NaOH), and gelatinized at 37 °C for 20 min. Tissue recombinants were cultured in DMEM/10% FBS supplemented with 10^{-7} M DHT overnight, followed by transplantation under the kidney capsules of immunodeficient NOD.Cg-*Prkdc^{scid} Il2rg^{tm1Wjl}*/SzJ (NSG) mice (Jackson Laboratory). Grafts were collected after 8–12 weeks of growth for analysis, and extreme limiting dilution analysis was performed as described[50].

Histology and immunostaining

H&E staining was performed using standard protocols on 6 μm paraffin sections. Histological assessments were performed using a published classification of mouse PIN lesions[24].

For immunohistochemical staining, 6 μm paraffin sections were deparaffinized in xylene, followed by boiling in antigen unmasking solution (Vector Labs). Slides were blocked in 10% normal goat serum (NGS) (Vector Labs), and incubated with primary antibodies diluted in 10% NGS overnight at 4 °C. Secondary antibodies were obtained from Vectastain ABC kits (Vector Labs) and diluted 1:250. Signal was enhanced using the Vectastain ABC system and visualized with the NovaRed Substrate Kit (Vector Labs). Slides were counterstained with Harris Modified Hematoxylin (1:4 diluted in H_2O) (Fisher Scientific) and mounted with Clearmount (American MasterTech). H&E and immunohistochemical

staining were imaged using a Nikon Eclipse E800 microscope equipped with a Nikon DXM1200 digital camera.

Immunofluorescence staining was performed using 6 μm cryosections, which were incubated in 3% H_2O_2 and Antigen Unmasking Solution (Vector Labs), or culture plates for whole-mount staining of spheres fixed in 4% paraformaldehyde for 15 min. Samples were incubated with 10% NGS and primary antibodies diluted in 10% NGS overnight at 4 °C. Samples were then incubated with secondary antibodies (diluted 1:500 in PBST) labeled with Alexa Fluor 488, 555, or 647 (Invitrogen/Molecular Probes). Detection of Nkx3.1 was enhanced using tyramide amplification (Invitrogen/Molecular Probes) by incubation of slides with HRP-conjugated secondary antibody (1:100 dilution) (Invitrogen/Molecular Probes), followed by incubation with tyramide 555 for 6 min. Slides were mounted with VECTASHIELD mounting medium with DAPI (Vector Labs). Immunofluorescence staining was imaged using a Leica TCS SP5 spectral confocal microscope. All primary antibodies and dilutions used are listed in Supplementary Table S11.

Cell numbers were counted manually using confocal ×40 and ×63 photomicrographs. Statistical analyses were performed using a two-sample *t*-test, χ^2 test, or Fisher's Exact test as appropriate. At least three animals for each experiment or genotype were analyzed.

RNA sequencing

Total RNA from prostate tissues was isolated using the Nucleospin RNA II kit (Clontech). The quantity and quality of each sample was measured using an Agilent 2100 Bioanalyzer. RNA-seq analysis was performed by the Columbia Genomic Sequencing Core Facility. FPKM values for 22,310 genes were reported. Genes with missing values (count of 0) in more than 10% of samples were eliminated from the analysis. Missing values for the remaining genes were estimated by using impute.knn (impute package in R v2.11.1). The resulting dataset of 14,063 genes was normalized by using robust spline normalization (RSN) function of R-system v2.11.1 and was log transformed. Principal Components Analysis was performed on scaled data, where the data value was adjusted by subtracting its mean across all samples and dividing by its standard deviation, z=(x-mean)/std dev. Expression data are deposited in the Gene Expression Omnibus database under GSE39509.

Statistical data analysis

Differential expression was estimated using the Welch T-test (t.test function in R v2.11.1). To compare two distinct signatures we used Gene Set Enrichment Analysis (GSEA)[51], where p-value was estimated with 1,000 sample permutations. Fold-change analysis was performed on data regenerated by reverse log transformation. Kaplan-Meier curves were generated using the surv, survdiff, and survfit functions from package survival in R. C-statistics analysis was conducted using the concordance index function from the R survcomp package. The multivariate Cox proportional hazards model analysis with the Gleason score and "leading edge" genes was computed using a multiplicative and additive model for defining an integrated risk score model.

Comparing initiation and progression genes for luminal and basal tumors

To compare genes that are responsible for initiation in tumors of both origins, we compared the L1 vs. Lc signature with the B3 vs. Bc signature. The query signature was defined as a list of genes ranked by their differential expression in the B3 group (n=5 samples) compared to the Bc group (n=6 samples) and was divided into two tails: a positive tail containing top 200 overexpressed genes in B3 compared to Bc, and a negative tail with top 200 under-expressed genes in B3 compared to Bc. The target signature was defined between L1 (n=6 samples) and Lc (n=5 samples) groups (Supplementary Tables S3, S4). Statistical significance of the enrichment between the query signatures and the target signatures was computed separately for the positive and negative tails using GSEA. Analogous analyses were performed comparing L3 vs. L1 with B6 vs. B3 signatures (Supplementary Tables S5, S6).

Comparing a luminal vs basal origin mouse signature with a human lethality signature

The luminal versus basal mouse signature (Fig. 5e,f) was defined between L3 (n=6 samples) and B6 groups (n=6 samples). For comparison with human prostate cancer expression profiling data, mouse genes were mapped to their human orthologs using mouse-human orthologous relations from the Mouse Genome Informatics database (MGI, http://www.informatics.jax.org/). This resulted in reduction of 14,063 mouse genes and 6,144 human genes to the set of 4,629 genes common between the human and mouse platforms.

The human “lethality risk” signature was derived from the Swedish watchful-waiting dataset[25], consisting of 281 prostate cancer samples from the cohort recruited in Sweden between 1977 and 1999. We defined the high-risk (aggressive) group as patients that survived less than 12 months (n=6) and the low-risk (indolent) group as patients that survived for more than 192 months (n=12).

Genes differentially expressed between mouse L3 and B6 tumors, p-value 0.01, were compared to the human lethality signature using GSEA. (Genes that were differentially expressed between the mouse models and did not differentially change between L3 and Lc or between B6 and Bc were excluded.) We divided the mouse gene list into those corresponding to genes overexpressed in L3 relative to B6 tumors (192 genes), and those corresponding to genes overexpressed in B6 compared to L3 (187 genes). The statistical significance of enrichment between the mouse query genes and the human target signature was computed separately for the group up in L3 tumors and the group up in B6 tumors.

Defining tumor subtypes in human patients

We examined the expression of the 68 genes in the leading-edge luminal origin signature (LOLES) in 5 high-quality patient datasets in the Oncomine database[22, 26–29], and found 19 genes to be up-regulated in prostate cancer relative to benign prostate tissue in at least 3 of the 5 datasets (Supplementary Fig. S6o, (Supplementary Table S10 . We used the LOLES to stratify the 79 primary tumor samples from the Glinsky dataset[30] and the 131 primary tumor samples from the Taylor dataset[22]. When multiple probes mapped to a single gene, the probe with the highest coefficient of variation was selected to represent the gene. To differentiate between aggressive and non-aggressive cases of prostate cancer, an event for the survival

analysis was defined as a biochemical recurrence happening within five years. The LOLES classified these patients into high-risk and low-risk classes after k-means clustering with the kmeans function in R. The difference between these two classes was tested with respect to BCR survival time (Kaplan-Meier BCR-free survival curve) and the p-value of this difference was computed with a log-rank test. For analysis of the Swedish cohort[25], we avoided overfitting by excluding patients used to construct the high-risk or low-risk human lethality signature (6 high-risk and 12 low-risk samples), resulting in 263 samples.

Pathway analysis

Enrichment of the mouse L3 vs. B6 signature in human biological pathways was evaluated by GSEA using pathways collected in the REACTOME[52], KEGG[53], and BioCarta (http://www.biocarta.com/genes/allpathways.asp) databases. We mapped mouse genes to human orthologs using mouse-human orthologous relations from the Mouse Genome Informatics database (http://www.informatics.jax.org/), and selected those mapped genes that appeared in at least one biological pathway, resulting in 9,945 unique genes.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Jason Rock and Brigid Hogan (Duke University Medical School) for their generous gift of *CK5-CreER*^{T2} transgenic mice, Olivier Couronne and the Columbia University Genome Sequencing Facility for RNA-seq analysis, and Ron DePinho (MD Anderson Cancer Center) for assistance in obtaining the Glinsky dataset. This work was supported by a post-doctoral fellowship from the DOD Prostate Cancer Research Program (Z.A.W.), a post-doctoral Computing Innovation Fellowship (CIFellow) from NSF, CRA, and CCC (A.M.), and by grants from the National Institutes of Health (C.A.-S., A.C., and M.M.S.).

References

1. Visvader JE. Cells of origin in cancer. Nature. 2011; 469:314–322. [PubMed: 21248838]
2. Shen MM, Abate-Shen C. Molecular genetics of prostate cancer: new prospects for old challenges. Genes Dev. 2010; 24:1967–2000. [PubMed: 20844012]
3. Goldstein AS, Stoyanova T, Witte ON. Primitive origins of prostate cancer: In vivo evidence for prostate-regenerating cells and prostate cancer-initiating cells. Mol Oncol. 2010; 4:385–396. [PubMed: 20688584]
4. Moscatelli D, Wilson EL. PINing down the origin of prostate cancer. Sci Transl Med. 2010; 2 43ps38.
5. Wang ZA, Shen MM. Revisiting the concept of cancer stem cells in prostate cancer. Oncogene. 2011; 30:1261–1271. [PubMed: 21119602]
6. De Marzo AM, Nelson WG, Bieberich CJ, Yegnasubramanian S. Prostate cancer: New answers prompt new questions regarding cell of origin. Nat Rev Urol. 2010; 7:650–652. [PubMed: 21139640]
7. Taylor RA, Toivanen R, Risbridger GP. Stem cells in prostate cancer: treating the root of the problem. Endocr Relat Cancer. 2010; 17:R273–R285. [PubMed: 20660571]
8. Lawson DA, Xin L, Lukacs RU, Cheng D, Witte ON. Isolation and functional characterization of murine prostate stem cells. Proc Natl Acad Sci USA. 2007; 104:181–186. [PubMed: 17185413]
9. Richardson GD, et al. CD133, a novel marker for human prostatic epithelial stem cells. J Cell Sci. 2004; 117:3539–3545. [PubMed: 15226377]

10. Burger PE, et al. Sca-1 expression identifies stem cells in the proximal region of prostatic ducts with high capacity to reconstitute prostatic tissue. Proc Natl Acad Sci USA. 2005; 102:7180–7185. [PubMed: 15899981]
11. Xin L, Lawson DA, Witte ON. The Sca-1 cell surface marker enriches for a prostate-regenerating cell subpopulation that can initiate prostate tumorigenesis. Proc Natl Acad Sci USA. 2005; 102:6942–6947. [PubMed: 15860580]
12. Goldstein AS, et al. Trop2 identifies a subpopulation of murine and human prostate basal cells with stem cell characteristics. Proc Natl Acad Sci USA. 2008; 105:20882–20887. [PubMed: 19088204]
13. Goldstein AS, Huang J, Guo C, Garraway IP, Witte ON. Identification of a cell of origin for human prostate cancer. Science. 2010; 329:568–571. [PubMed: 20671189]
14. Wang X, et al. A luminal epithelial stem cell that is a cell of origin for prostate cancer. Nature. 2009; 461:495–500. [PubMed: 19741607]
15. Shibata M, Shen MM. The roots of cancer: Stem cells and the basis for tumor heterogeneity. Bioessays. 2012 epub ahead of print.
16. Visvader JE. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. Genes Dev. 2009; 23:2563–2577. [PubMed: 19933147]
17. Lawson DA, et al. Basal epithelial stem cells are efficient targets for prostate cancer initiation. Proc Natl Acad Sci USA. 2010; 107:2610–2615. [PubMed: 20133806]
18. Choi N, Zhang B, Zhang L, Ittmann M, Xin L. Adult murine prostate basal and luminal cells are self-sustained lineages that can both serve as targets for prostate cancer initiation. Cancer Cell. 2012; 21:253–265. [PubMed: 22340597]
19. Rock JR, et al. Basal cells as stem cells of the mouse trachea and human airway epithelium. Proc Natl Acad Sci USA. 2009; 106:12771–12775. [PubMed: 19625615]
20. Srinivas S, et al. Cre reporter strains produced by targeted insertion of EYFP and ECFP into the ROSA26 locus. BMC Dev Biol. 2001; 1:4. [PubMed: 11299042]
21. Xin L, Lukacs RU, Lawson DA, Cheng D, Witte ON. Self-renewal and multilineage differentiation in vitro from murine prostate stem cells. Stem Cells. 2007; 25:2760–2769. [PubMed: 17641240]
22. Taylor BS, et al. Integrative genomic profiling of human prostate cancer. Cancer Cell. 2010; 18:11–22. [PubMed: 20579941]
23. Berger MF, et al. The genomic complexity of primary human prostate cancer. Nature. 2011; 470:214–220. [PubMed: 21307934]
24. Park JH, et al. Prostatic intraepithelial neoplasia in genetically engineered mice. Am J Pathol. 2002; 161:727–735. [PubMed: 12163397]
25. Sboner A, et al. Molecular sampling of prostate cancer: a dilemma for predicting disease progression. BMC medical genomics. 2010; 3:8. [PubMed: 20233430]
26. Lapointe J, et al. Gene expression profiling identifies clinically relevant subtypes of prostate cancer. Proc Natl Acad Sci USA. 2004; 101:811–816. [PubMed: 14711987]
27. Tomlins SA, et al. Integrative molecular concept modeling of prostate cancer progression. Nat Genet. 2007; 39:41–51. [PubMed: 17173048]
28. Singh D, et al. Gene expression correlates of clinical prostate cancer behavior. Cancer Cell. 2002; 1:203–209. [PubMed: 12086878]
29. Yu YP, et al. Gene expression alterations in prostate cancer predicting tumor aggression and preceding development of malignancy. J Clin Oncol. 2004; 22:2790–2799. [PubMed: 15254046]
30. Glinsky GV, Glinskii AB, Stephenson AJ, Hoffman RM, Gerald WL. Gene expression profiling predicts clinical outcome of prostate cancer. J Clin Invest. 2004; 113:913–923. [PubMed: 15067324]
31. Martin NE, Mucci LA, Loda M, Depinho RA. Prognostic determinants in prostate cancer. Cancer J. 2011; 17:429–437. [PubMed: 22157287]
32. Snippert HJ, Clevers H. Tracking adult stem cells. EMBO Rep. 2011; 12:113–122. [PubMed: 21252944]
33. Taylor RA, et al. Lineage enforcement by inductive mesenchyme on adult epithelial stem cells across developmental germ layers. Stem Cells. 2009; 27:3032–3042. [PubMed: 19862839]

34. Cunha GR, et al. The endocrinology and developmental biology of the prostate. Endocrine Rev. 1987; 8:338–362. [PubMed: 3308446]
35. Van Keymeulen A, et al. Distinct stem cells contribute to mammary gland development and maintenance. Nature. 2011; 479:189–193. [PubMed: 21983963]
36. Lu CP, et al. Identification of stem cell populations in sweat glands and ducts reveals roles in homeostasis and wound repair. Cell. 2012; 150:136–150. [PubMed: 22770217]
37. Liu J, et al. Regenerated luminal epithelial cells are derived from preexisting luminal epithelial cells in adult mouse prostate. Mol Endocrinol. 2011; 25:1849–1857. [PubMed: 21940754]
38. Ousset M, et al. Multipotent and unipotent progenitors contribute to prostate postnatal development. Nat Cell Biol. 2012; 14:1131–1138. [PubMed: 23064263]
39. Blackwood JK, et al. In situ lineage tracking of human prostatic epithelial stem cell fate reveals a common clonal origin for basal and luminal cells. J Pathol. 2011; 225:181–188. [PubMed: 21898876]
40. Gaisa NT, et al. Clonal architecture of human prostatic epithelium in benign and malignant conditions. J Pathol. 2011; 225:172–180. [PubMed: 21898875]
41. Doupe DP, et al. A single progenitor population switches behavior to maintain and repair esophageal epithelium. Science. 2012; 337:1091–1093. [PubMed: 22821983]
42. Doupe DP, Klein AM, Simons BD, Jones PH. The ordered architecture of murine ear epidermis is maintained by progenitor cells with random fate. Dev Cell. 2010; 18:317–323. [PubMed: 20159601]
43. Markert EK, Mizuno H, Vazquez A, Levine AJ. Molecular classification of prostate cancer using curated expression signatures. Proc Natl Acad Sci USA. 2011; 108:21276–21281. [PubMed: 22123976]
44. Lapointe J, et al. Genomic profiling reveals alternative genetic pathways of prostate tumorigenesis. Cancer Res. 2007; 67:8504–8510. [PubMed: 17875689]
45. Rubin MA, Maher CA, Chinnaiyan AM. Common gene rearrangements in prostate cancer. J Clin Oncol. 2011; 29:3659–3668. [PubMed: 21859993]
46. Alizadeh AA, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000; 403:503–511. [PubMed: 10676951]
47. Gibson P, et al. Subtypes of medulloblastoma have distinct developmental origins. Nature. 2010; 468:1095–1099. [PubMed: 21150899]
48. Collisson EA, et al. Subtypes of pancreatic ductal adenocarcinoma and their differing responses to therapy. Nat Med. 2011; 17:500–503. [PubMed: 21460848]
49. Banach-Petrosky W, et al. Prolonged exposure to reduced levels of androgen accelerates prostate cancer progression in Nkx3.1; Pten mutant mice. Cancer Res. 2007; 67:9089–9096. [PubMed: 17909013]
50. Hu Y, Smyth GK. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. Journal of immunological methods. 2009; 347:70–78. [PubMed: 19567251]
51. Subramanian A, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci USA. 2005; 102:15545–15550. [PubMed: 16199517]
52. Croft D, et al. Reactome: a database of reactions, pathways and biological processes. Nucleic Acids Res. 2011; 39:D691–D697. [PubMed: 21067998]
53. Ogata H, et al. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 1999; 27:29–34. [PubMed: 9847135]

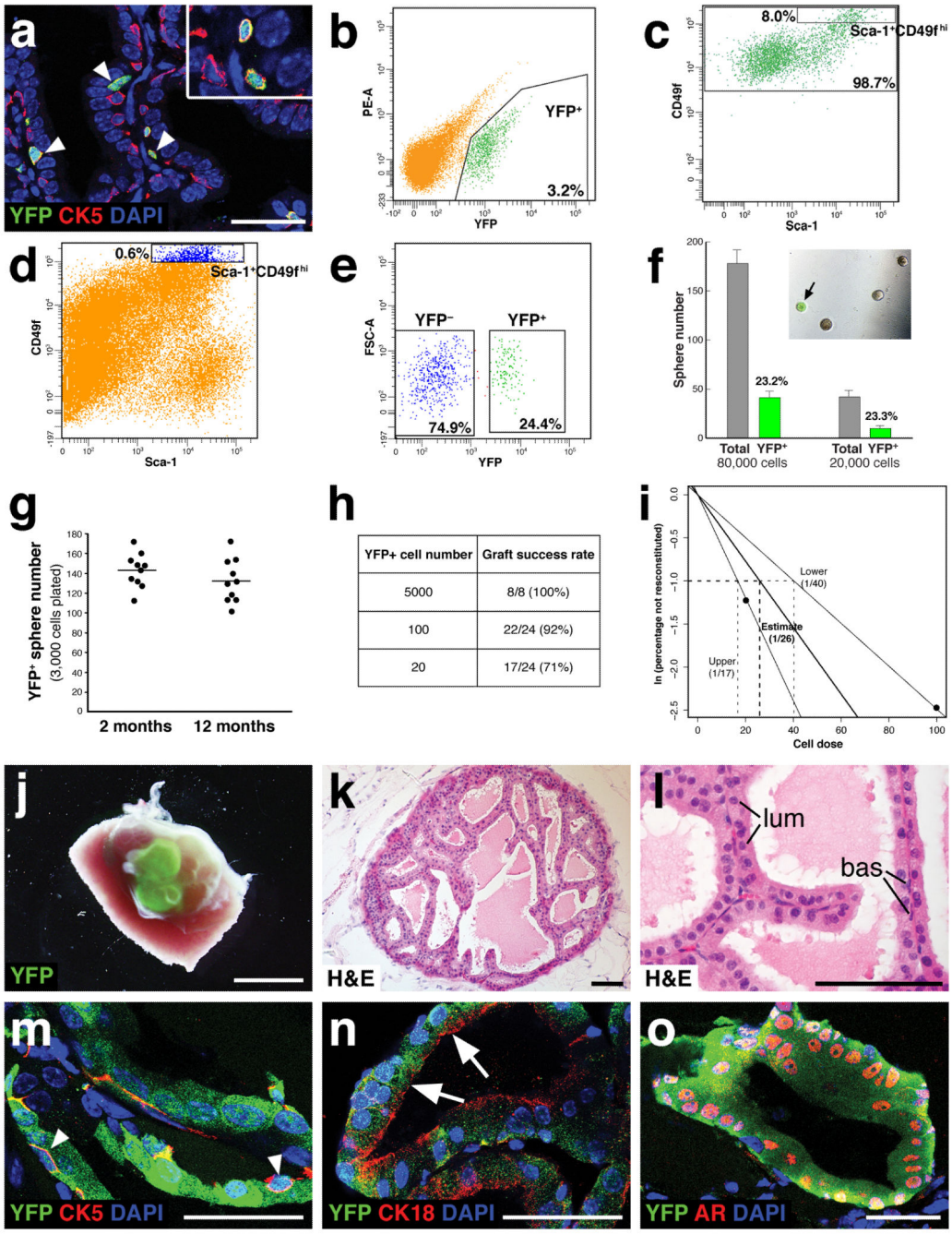

Figure 1.

High frequency of prostate basal stem/progenitor cells in sphere formation and tissue reconstitution assays. For all analyses, tamoxifen-induced *CK5-CreER^{T2}; R26R-YFP/+* mice were analyzed at 14 days after tamoxifen treatment. (**a**) Immunofluorescence staining showing co-localization of YFP with CK5 in basal cells (arrowheads) of the anterior prostate. (**b**) Purification of YFP^{+} basal cells from dissociated prostate tissue by flow cytometry. (**c**) Flow-sorting of YFP^{+} cells shows that 98.7% are CD49f^{+}, and 8.0% are Lin^{-}Sca-1^{+}CD49f^{hi} cells. (**d,e**) Flow sorting of Lin^{-}Sca-1^{+}CD49f^{hi} cells (panel d, 0.6% of

total Lin^- cells) shows that 24.4% are YFP^+ (**e**). (**f**) Quantitation of sphere formation from 80,000 or 20,000 dissociated prostate cells, showing the number of total spheres as well as YFP^+ spheres. Each experiment was performed six times, using two replicates each from three independent mice; error bars correspond to standard deviation and show variability between the six samples. Inset shows epifluorescence detection of YFP expression in spheres (arrow). (**g**) Quantitation of sphere formation from 3,000 dissociated YFP^+ cells isolated from *CK5-CreER^{T2}; R26R-YFP/+* mice at 2 months or 12 months of age. The frequencies at these two stages are not statistically different by two sample t-test. (**h**) Serial dilution analysis of purified YFP^+ cells in assays of prostate duct formation in renal grafts. (**i**) Extreme limiting dilution analysis of data in (**h**). (**j**) YFP fluorescence of a renal graft attached to a portion of kidney tissue. (**k,l**) Hematoxylin-eosin staining of a tissue section from a renal graft generated from purified YFP^+ basal cells; luminal (lum) and basal (bas) cells are indicated (**l**). (**m-o**) Analysis of YFP together with CK5 expression in basal cells (arrowheads, **m**), CK18 in luminal cells (arrows, **n**), and AR (**o**) in renal grafts. Scale bars in **a,m-o** correspond to 50 microns, in **k,l** to 100 microns, and in **j** to 1 mm.

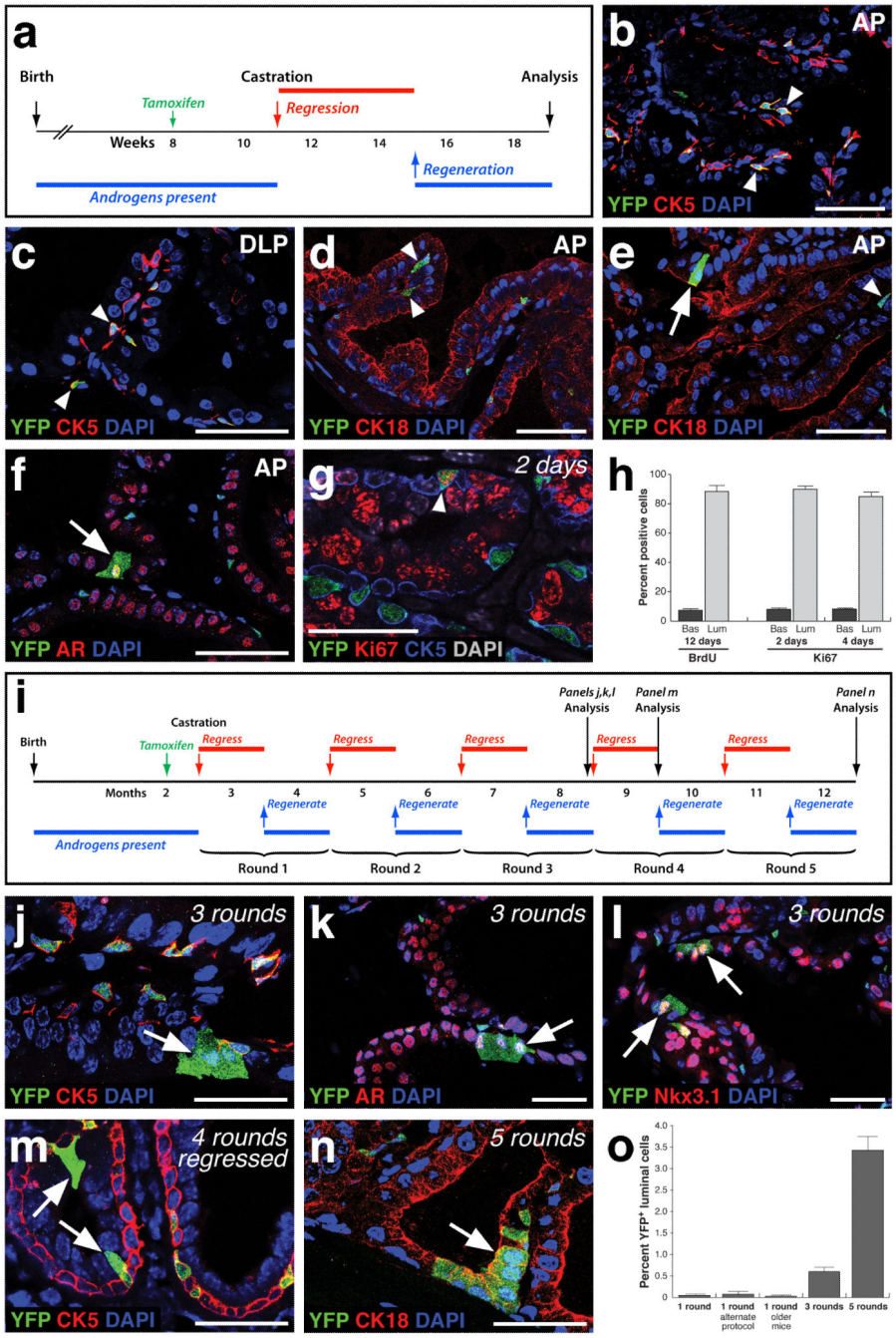

Figure 2.

Detection of rare bipotential basal progenitors during prostate regeneration *in vivo*. (**a**) Lineage-tracing strategy during prostate regeneration of *CK5-CreER^{T2}; R26R-YFP/+* mice. (**b,c**) Co-localization of YFP with CK5-expressing basal cells (arrowheads) in regenerated *CK5-CreER^{T2}; R26R-YFP/+* anterior prostate (**b**) and dorsolateral prostate (**c**). (**d-f**) Most YFP$^+$ cells (arrowheads) do not express the luminal marker CK18, although rare YFP$^+$CK18$^+$ (arrow, **e**) and YFP$^+$AR$^+$ (arrow, **f**) luminal cells can be detected after one round of regeneration. (**g**) Ki67 immunostaining at two days after androgen administration

shows that most luminal and some basal cells (arrowhead) undergo proliferation. (**h**) Quantitation of cell proliferation assays during regeneration, showing that 7.4% of basal cells (n=510/6,929) and 88.7% of luminal cells (n=6,708/7,565) were $BrdU^+$ after 12 days of incorporation; 8.0% of basal cells (n=426/5,326) and 90.1% of luminal cells (n=6,945/7,709) were $Ki67^+$ at 2 days of regeneration; and 8.1% of basal cells (n=380/4,700) and 85.1% of luminal cells (n=7,094/8,333) were $Ki67^+$ at 4 days of regeneration; 3 animals were analyzed for each experiment. See Supplementary Fig. S3 and Supplementary Table S1 for additional data. (**i**) Lineage-tracing strategy during serial regression and regeneration. (**j-l,n**) YFP^+ luminal cells (arrows) that co-express AR (**k**), Nkx3.1 (**l**), and CK18 (**n**) are more frequently observed after three rounds (**j-l**) and five rounds (**n**) of serial regeneration. (**m**) Castration-resistant luminal cells can be detected after four rounds of regression. (**o**) The frequency of luminal cells among total YFP^+ cells during regeneration in anterior prostate is 0.04% (n=5/11,427, 3 animals) after 1 round, 0.07% (n=13/18,025, 5 animals) after 1 round using an alternate protocol, 0.03% (n=3/10,249, 3 animals) after 1 round in 12-month old mice, 0.6% (n=56/9,129, 3 animals) after 3 rounds, and 3.4% (n=303/8,875, 3 animals) after 5 rounds. The 1 round frequencies are not statistically different, while $p<0.0001$ for frequencies of different rounds by χ^2 test. See Supplementary Fig. S4 and Supplementary Table S1 for additional data. Abbreviations: AP, anterior prostate; DLP, dorsolateral prostate. Scale bars in **b-g,j-n** correspond to 50 microns; error bars in **h,o** correspond to standard deviation and show variability between animals.

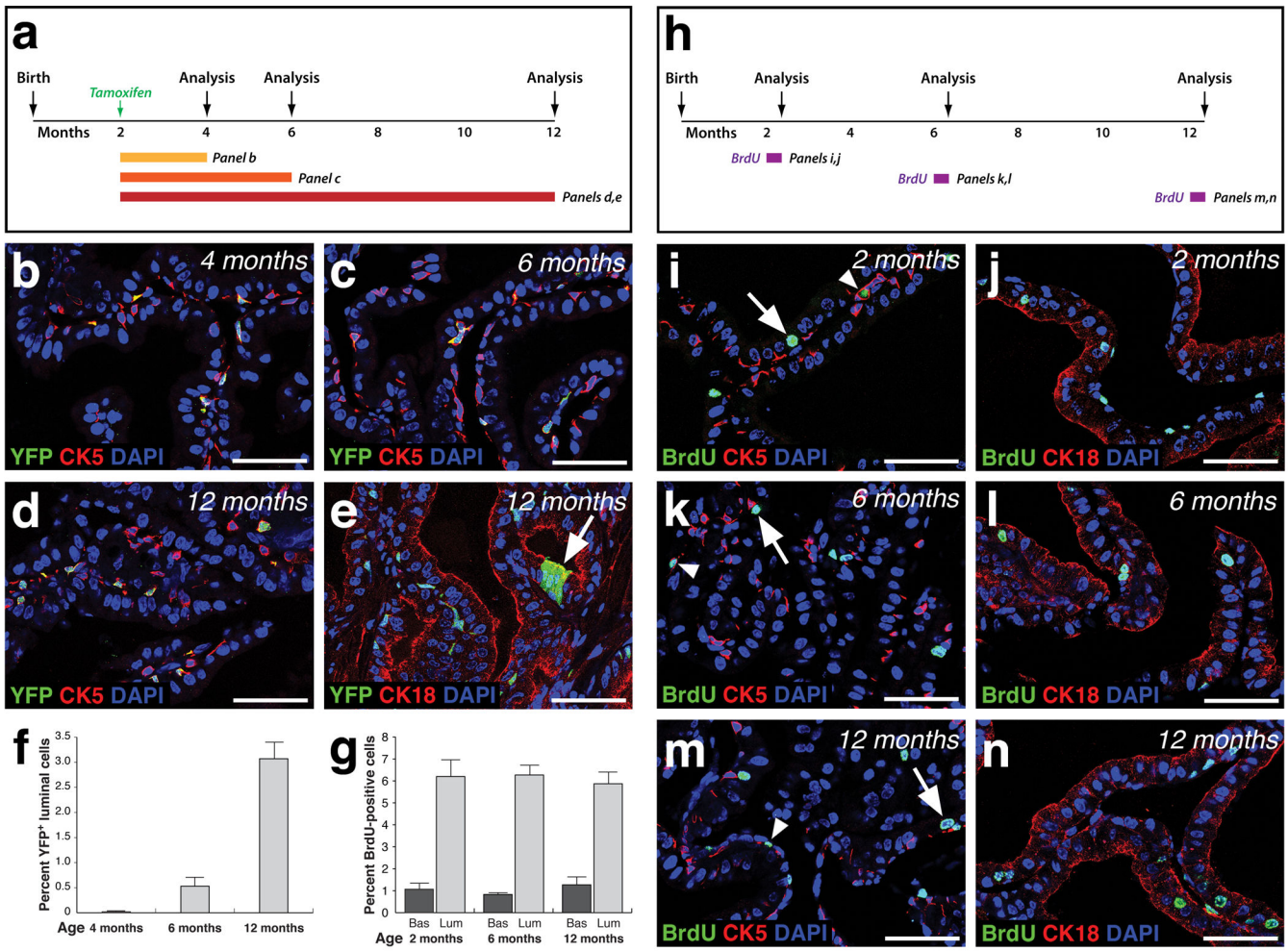

Figure 3.

Detection of rare bipotential basal progenitors during prostate homeostasis. (**a**) Time course of lineage-tracing analysis in hormonally intact *CK5-CreER^{T2}; R26R-YFP/+* mice. (**b-d**) Co-localization of YFP and CK5 in prostate basal cells in mice at 4 months (**b**), 6 months (**c**), and 12 months (**d**) of age. (**e**) Detection of YFP^{+}CK18^{+} luminal cells (arrow) at 12 months of age. (**f**) The frequency of luminal cells among total YFP^{+} cells during homeostasis is 0.02% (n=2/8,848) at 4 months, 0.5% (n=57/10,572) at 6 months, and 3.0% (n=227/7,638) at 12 months; 3 animals were analyzed at each time point. $p<0.0001$ for frequencies at different time points by χ^2 test. (**g**) Graphical summary of BrdU incorporation analyses during homeostasis; 3 animals were analyzed for each experiment. BrdU incorporation frequencies at different time points are not statistically different by χ^2 test. See Supplementary Table S1 for additional data. (**h**) Strategy for analyses of cell proliferation at three different ages in wild-type C57BL/6 mice. At 2 months, 6 months, or 12 months of age, BrdU was administered for 12 days followed by analysis. (**i-n**) Analysis of co-localization of BrdU immunostaining with CK5 (**i,k,m**) or CK18 (**j,l,n**); arrowheads indicate BrdU-positive basal cells, and arrows indicate BrdU-positive luminal cells. Scale bars in **b-**

e,i-n correspond to 50 microns; error bars in **f,g** correspond to standard deviation and show variability between animals.

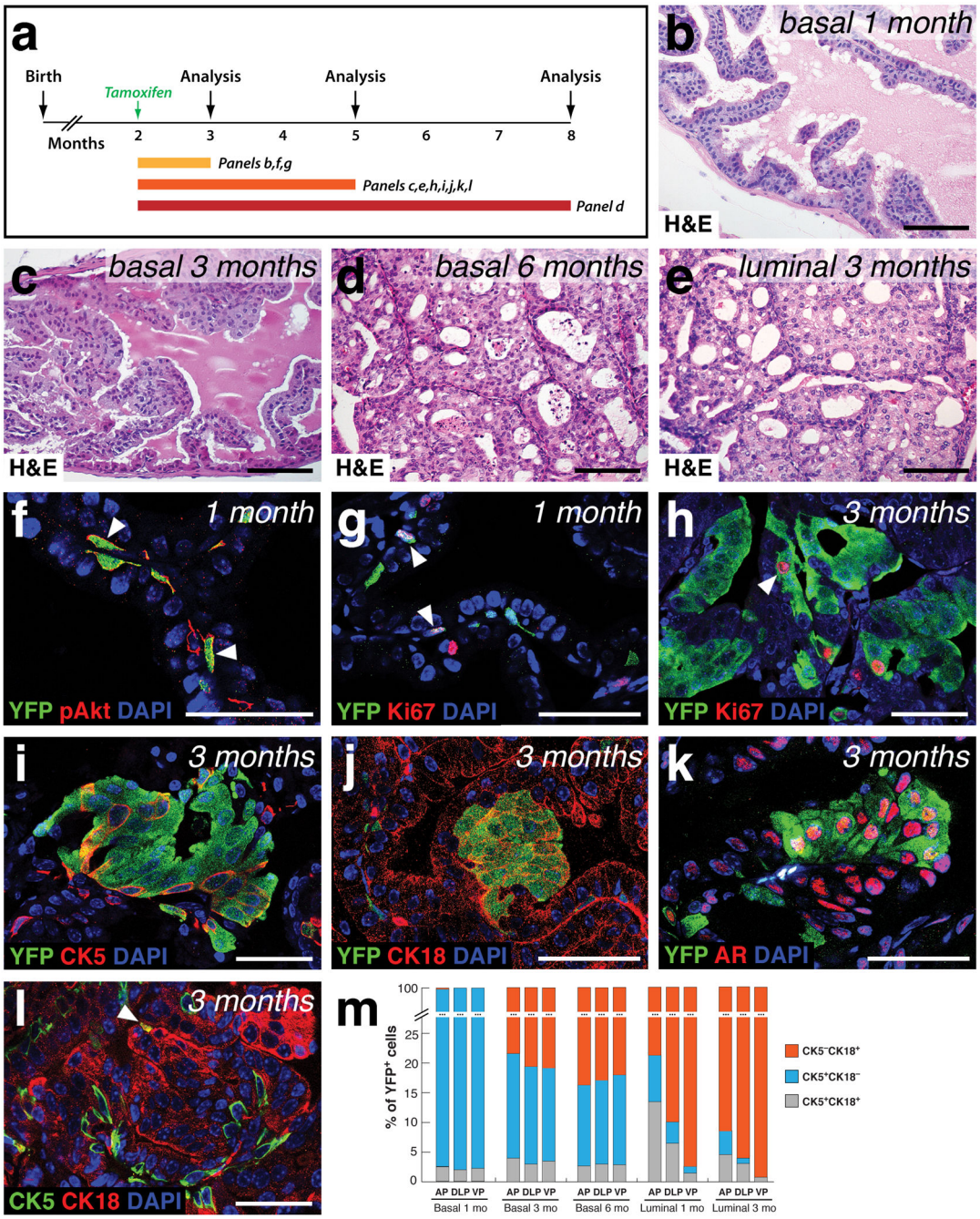

Figure 4.

Basal cells are a cell type of origin for prostate tumors. (**a**) Time course for tumor formation in hormonally-intact *CK5-CreER^{T2}; Pten$^{flox/flox}$; R26R-YFP/+* mice. (**b-d**) Hematoxylin-eosin staining of anterior prostates showing slight epithelial hyperplasia at 1 month after induction (**b**), low-grade PIN at 3 months after induction (**c**), and high-grade PIN at 6 months after induction (**d**). (**e**) High-grade PIN in *Nkx3.1$^{CreERT2/+}$; Pten$^{flox/flox}$; R26R-YFP/+* anterior prostate at 3 months after induction. (**f-l**) Marker analysis of PIN lesions in *CK5-CreER^{T2}; Pten$^{flox/flox}$; R26R-YFP/+* anterior prostate. (**f**) Phosphorylated Akt (pAkt)

can be detected in basal cells at 1 month after induction. (**g,h**) Ki67 immunoreactivity can be detected in basal cells at one month after induction, prior to PIN lesion formation (**g**), as well as at three months after induction (**h**). (**i,j,k**) Most transformed cells at three months after induction do not express CK5 (**i**), but instead express CK18 (**j**) and AR (**k**). (**l**) $CK5^{+}CK18^{+}$ intermediate cells (arrowhead) can be detected in PIN lesions at three months after induction. (**m**) Quantitation of basal ($CK5^{+}CK18^{-}$), luminal ($CK5^{-}CK18^{+}$), and intermediate cells ($CK5^{+}CK18^{+}$) in YFP^{+} prostate cells of *CK5-CreER^{T2}; Pten$^{flox/flox}$; R26R-YFP/+* and *Nkx3.1$^{CreERT2/+}$; Pten$^{flox/flox}$; R26R-YFP/+* mice at the indicated times after induction. Scale bars in **b-e** correspond to 100 microns, and in **f-l** to 50 microns.

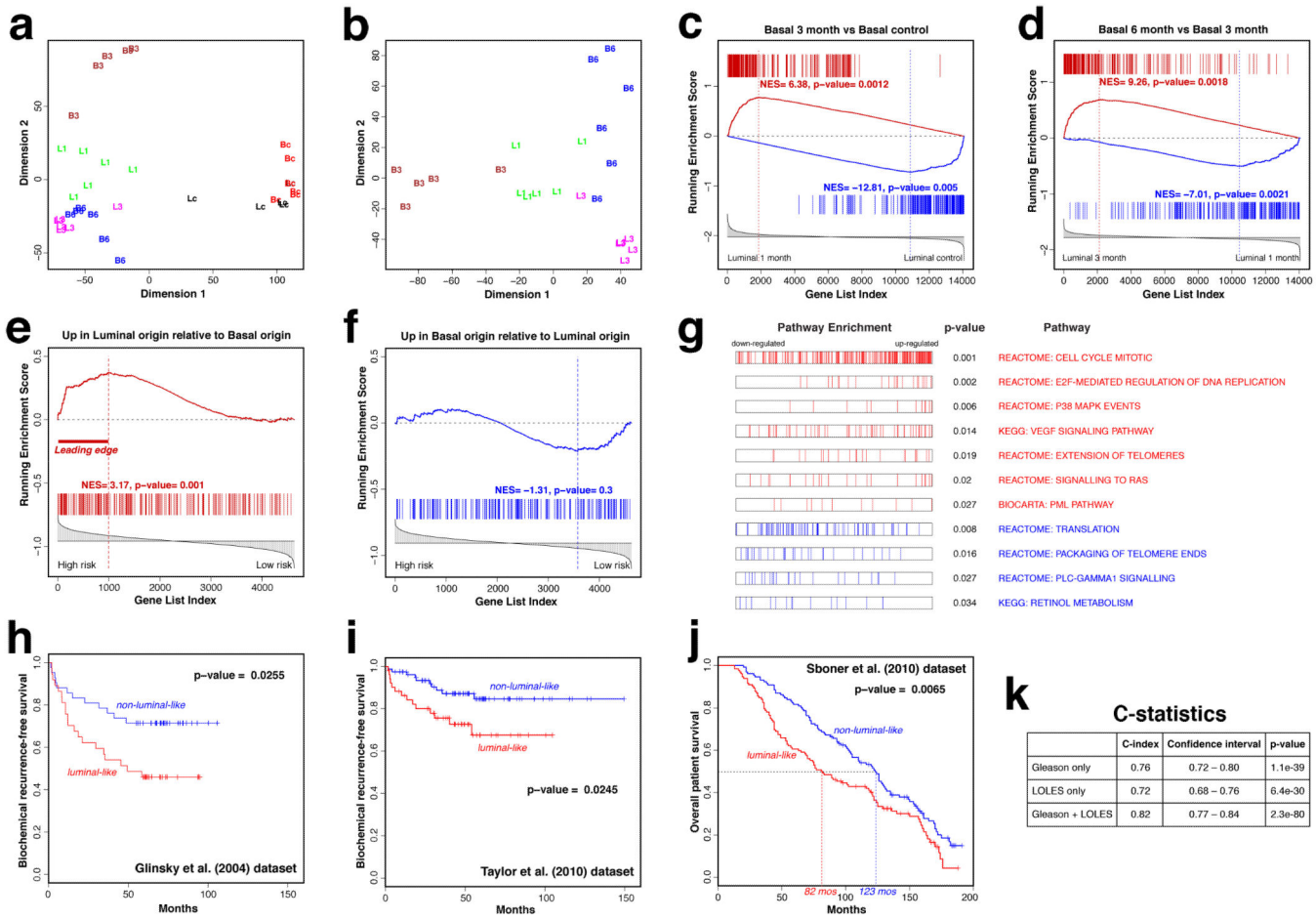

Figure 5.

A luminal origin gene signature that is prognostic for human prostate cancer outcome. (**a**) Scatter-plot of the two main components from a Principal Components Analysis based on 14,063 genes, capturing 55% (dimension 1) and 21% (dimension 2) of the data variability. (**b**) Scatter-plot without the control samples. Abbreviations: Bc, basal origin control; B3, basal origin 3 months post-induction; B6, basal origin 6 months post-induction; Lc, luminal origin control; L1, luminal origin 1 month post-induction; L3, luminal origin 3 months post-induction. (**c**) GSEA comparison of basal origin “initiation” signature (B3 relative to Bc) to luminal “initiation” signature (L1 relative to Lc) shows strong enrichment in both directions. (**d**) GSEA comparison of basal origin “progression” signature (B6 relative to B3) to luminal “progression” signature (L3 relative to L1) also shows strong enrichment in both directions. (**e,f**) GSEA shows that genes up-regulated in the L3 versus B6 gene signature are strongly enriched in a human signature corresponding to lethality due to prostate cancer (**e**), but not the converse (**f**). (**g**) GSEA shows biological pathways significantly enriched in the L3 versus B6 gene signature; p-value is estimated using 1,000 sample permutations. Pathways in red are up-regulated in luminal origin tumors while pathways in blue are up-regulated in basal origin tumors. (**h,i**) Kaplan-Meier analysis shows that the LOLES (corresponding to the 68 genes to the left of the dashed line in **e**) stratifies patients from two independent cohorts into groups with different rates of biochemical recurrence (red curve, “luminal-like”

group, 37 patients in **h**, 52 patients in **i**; blue curve, “non-luminal-like” group, 42 patients in **h**, 79 patients in **i**). (**j**) Kaplan-Meier analysis shows that the LOLES stratifies patients from the Swedish “watchful waiting” cohort into a “luminal-like” group (red curve, 132 patients) and a “non-luminal-like” group (blue curve, 131 patients) with different survival outcomes (82 months versus 123 months at 50% survival). (**k**) C-statistics analysis shows that the LOLES improves the prognostic value of Gleason score in the Swedish cohort from 0.76 to 0.82, with the 95% confidence intervals and p-values shown.

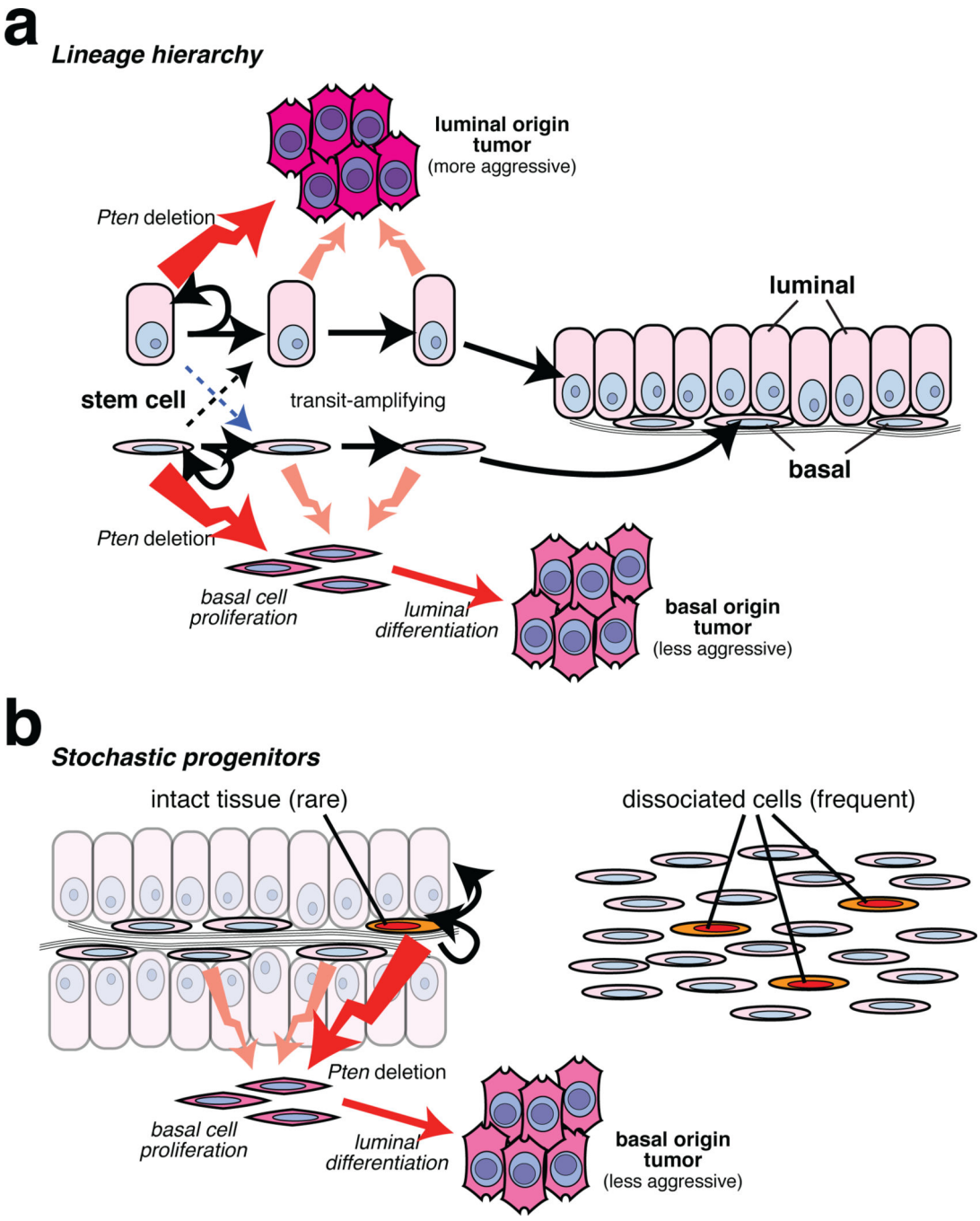

Figure 6.

Two models for prostate epithelial lineage relationships and cell of origin for cancer. (**a**) In a conventional lineage hierarchy model, luminal and basal lineages are independently maintained by largely unipotent stem/progenitor cells in the normal adult prostate epithelium. However, luminal and basal progenitors can generate the other cell type during prostate regeneration and tissue homeostasis (dashed lines); in the case of luminal stem/progenitor cells, it remains unclear whether such bipotentiality (blue dashed line) is displayed only by CARNs in the regressed state. In the case of the basal lineage, bipotential

stem cells are relatively rare (approximately 0.05%), while basal cells that can display stem cell properties in sphere formation and tissue reconstitution assays are more common (approximately 4%), and perhaps might correspond to transit-amplifying cells. Oncogenic transformation of either luminal or basal cells by inactivation of *Pten* results in tumors with histologically similar luminal phenotypes, but tumors arising from basal cells first undergo basal cell proliferation and subsequently luminal differentiation. Tumors may arise from stem cells (dark red jagged arrows) or may also be derived from more differentiated cell types (light red jagged arrows). (**b**) In a stochastic progenitor model, basal cells within an intact prostate epithelium randomly display stem/progenitor properties at very low frequencies (orange), giving rise to luminal cells and being capable of self-renewal. (Luminal cells could conceivably follow a similar stochastic progenitor model, but this is not shown.) After tissue dissociation, however, the probability of such random basal stem/progenitor cells may be greatly increased. Oncogenic transformation of normal basal cells and/or stochastic basal progenitors leads to luminal differentiation and tumor formation; however, it is unlikely that stochastic progenitors represent the sole cell of origin following *Pten* deletion, given the rarity of these progenitors versus the frequency of observed PIN lesions.